

Disruption of the neurexin 1 gene is associated with schizophrenia

Dan Rujescu^{1#}, Andres Ingason^{2,3#}, Sven Cichon^{4,5}, Olli P.H. Pietiläinen⁶, Michael R Barnes⁷, Timothea Touloupoulou⁸, Marco Picchioni⁸, Evangelos Vassos⁸, Ulrich Ettinger⁸, Elvira Bramon⁸, Robin Murray⁸, Mirella Ruggeri⁹, Sarah Tosato⁹, Chiara Bonetto⁹, Stacy Steinberg², Engilbert Sigurdsson¹⁰, Thordur Sigmundsson¹⁰, Hannes Petursson¹⁰, Arnaldur Gylfason², Pall I Olason², Gudmundur Hardarsson², Gudrun A Jonsdottir², Omar Gustafsson², Ragnheidur Fossdal², Ina Giegling¹, Hans-Jürgen Möller¹, Annette Hartmann¹, Per Hoffmann⁴, Caroline Crombie¹¹, Gillian Fraser¹¹, Nicholas Walker¹², Jouko Lonnqvist¹³, Jaana Suvisaari¹³, Annamaria Tuulio-Henriksson¹³, Ole A. Andreassen^{14,15}, Srdjan Djurovic^{14,15}, Thomas Hansen³, Thomas Werge³, Ingrid Melle^{14,15}, Lambertus A. Kiemeny¹⁶, Barbara Franke¹⁷, Jacobine E. Buizer-Voskamp^{18,19}, Roel A. Ophoff^{19,20}, GROUP Investigators²¹, Marcella Rietschel²², Markus M. Nöthen^{4,5}, Kari Stefansson¹, Leena Peltonen^{6,15,23,24}, David St Clair¹¹, Hreinn Stefansson², David A Collier^{*25}

#Note of equal contribution

1. Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Munich, Germany.
2. deCODE genetics, Reykjavík, Iceland
3. Research Institute of Biological Psychiatry Mental Health Centre Sct. Hans, Copenhagen University Hospital, Roskilde, Denmark
4. Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany.
5. Institute of Human Genetics, University of Bonn, Bonn, Germany
6. Department for Molecular Medicine, National Public Health Institute, Helsinki, Finland
7. Computational Biology, GlaxoSmithKline Pharmaceuticals, Harlow, Essex, UK
8. Division of Psychological Medicine, Institute of Psychiatry, King's College, London, UK
9. Section of Psychiatry and Clinical Psychology, University of Verona, Verona, Italy
10. Department of Psychiatry, National University Hospital, Reykjavík, Iceland
11. Department of Mental Health, University of Aberdeen, Aberdeen, Scotland
12. Ravenscraig Hospital, Greenock, Scotland
13. Department of Mental Health and Addiction, National Public Health Institute, Helsinki, Finland
14. Institute of Psychiatry, University of Oslo, Oslo, Norway

15. Department of Medical Genetics, Ulleval University Hospital, Oslo, Norway
16. Dept of Epidemiology & Biostatistics (133 EPIB) / Dept of Urology (659 URO) Radboud University Nijmegen Medical Centre Nijmegen, The Netherlands
17. Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
18. Rudolf Magnus Institute of Neuroscience, Department of Psychiatry, University Medical Centre Utrecht, Utrecht, The Netherlands
19. Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands
20. UCLA Center for Neurobehavioral Genetics, Los Angeles
21. Dept of Psychiatry, University Medical Centre Utrecht, Utrecht, The Netherlands[^]
22. Dept. of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health Mannheim, University of Heidelberg, Mannheim, Germany
23. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
24. The Broad Institute, Cambridge, MA, USA
25. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College, London, UK

GROUP investigators include; Don Linszen, René S. Kahn, Jim van Os, Durk Wiersma, Richard Bruggeman, Wiepke, Inez Germeys, Lieuwe de Haan and Lydia Krabbendam.

*Corresponding author: David A Collier PhD Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry (King's College London) De Crespigny Park, Denmark Hill London SE5 8AF
tel: +44 20 7848 0631 fax: +44 20 7848 0802 E-Mail: d.collier@iop.kcl.ac.uk

ABSTRACT

Deletions within the neurexin 1 gene (NRXN1; 2p16.3) are associated with autism and have also been reported in two families with schizophrenia. We examined NRXN1, and the closely related NRXN2 and NRXN3 genes, for copy number variants (CNVs) in 2,977 schizophrenia patients and 33,746 controls from seven European populations (Iceland, Finland, Norway, Germany, The Netherlands, Italy and UK) using microarray data. We found 66 deletions and 5 duplications in NRXN1, including a *de novo* deletion: 12 deletions and 2 duplications occurred in schizophrenia cases (0.47%) compared to 49 and 3 (0.15%) in controls. There was no common breakpoint and the CNVs varied from 18kb to 420kb. No CNVs were found in NRXN2 or NRXN3. We performed a Cochran-Mantel-Haenszel exact test to estimate association between all CNVs and schizophrenia ($P = 0.13$; $OR = 1.73$; 95%CI 0.81-3.50). Because the penetrance of NRXN1 CNVs may vary according to the level of functional impact on the gene, we next restricted the association analysis to CNVs that disrupt exons (0.17% of cases and 0.020% of controls). These were significantly associated with a high odds ratio ($P = 0.0027$; $OR = 8.97$, 95% CI 1.8-51.9). We conclude that NRXN1 deletions affecting exons confer risk of schizophrenia.

INTRODUCTION

Copy number variants (CNVs) are emerging as an important genomic cause of disease, particularly autism where large *de novo* CNVs may account for up to 10% of cases (1-3). These CNVs appear to be highly deleterious and may account for the dramatically reduced fecundity in autism; in line with autistic individuals, schizophrenics have fewer children on average than the general population and risk increases with paternal age (4,5). It also appears that the same CNVs can increase susceptibility to more than one psychiatric phenotype (6). Because of this, CNVs are also strong candidates for involvement in the aetiology of schizophrenia.

The role of CNVs in schizophrenia is less well documented, but there are classical examples such as the translocation disrupting the DISC1 gene in a large Scottish pedigree (7) and velo-cardio-facial syndrome (VCFS), caused by a large deletion at 22q11 which is associated with psychosis in about 30% of cases (8). Several recent papers have also indicated that there might be further schizophrenia associated CNVs, especially *de novo* deletions and duplications (9-14).

Both Kirov et al. (9) and Walsh et al. (11) investigated individuals with schizophrenia in order to identify microdeletions and microduplications >100kb potentially associated with the disease. Each study found a deletion disrupting the neurexin 1 gene (*NRXN1*) at 2p16.3 in one family, in a mother and two affected siblings, and identical twins concordant for child-onset schizophrenia respectively. However neither was able to determine the prevalence, penetrance or odds ratio of these CNVs because of their rarity and the small population size examined. CNVs in *NRXN1* have also previously been implicated in autism and mental retardation (6, 15-19). In the present study, we assessed the association of *NRXN1* CNVs with schizophrenia as part of a genome-wide scan in a sample of 2,977 schizophrenia patients and 33,746 controls from seven European populations (Iceland, Finland, Norway, Germany, The Netherlands, Italy and UK) using microarray data.

RESULTS

For the *NRXN1* locus on 2p16.3 we limited our search to include the region between Mb 49,90-51,50 (Human Genome build36), encompassing all exons of *NRXN1* and several hundred kilobases of the upstream sequence. For comparison, the deletions reported in schizophrenic patients by Kirov et al. (11) spans Mb 51,04-51,29 (Genome build36, 250Kb) and the deletion reported by Walsh et al. (13) Mb 50,023-50,137 (Genome build 36, 115kb). Similarly we examined the interval for deletions or duplications and tested for their enrichment in schizophrenia (see materials and methods section, CNV detection subsection).

We examined the region in 2,977 schizophrenia patients and 33,746 controls from seven European populations (Iceland, Finland, Norway, Germany, The Netherlands, Italy and UK) (supplementary table 1). We found 61 *NRXN1* deletions, including a *de novo* deletion, and 5 duplications in the combined sample (figure 1). The CNVs were found throughout the locus, and varied in size from 18 Kb to 420 Kb, with no common breakpoint, although some single breakpoints may be shared. We found no obvious sequence elements such as low copy repeats (LCRs) in the region that could mediate the rearrangements. In contrast there were no deletions or duplications found at the *NRXN2* (11p13.1) or *NRXN3* (14q31.1) loci, which were searched using an identical strategy.

We found 12 deletions and 2 duplications of *NRXN1* in cases (0.47%) compared to 49 and 3 (0.15%) in the controls, all of which encompassed part of the *NRXN1* gene or its promoter (figure 1). The most common deletion was found in 27 samples including one homozygote. One of these was schizophrenic, and two had a diagnosis of alcoholism; the remainder were controls (supplementary table 2). Haplotype background suggests that all carriers for the common deletion derive from a single founder. A further six deletions and one duplication were seen in two, three and four samples respectively while 24 deletions and one duplication were only observed once.

We performed Cochran-Mantel-Haenszel exact test to estimate association between all deletions and duplications and schizophrenia. This analysis gave suggestive results, $P = 0.13$

(OR=1.73; 95%CI 0.81-3.50) for all *NRXN1* CNVs in schizophrenia versus controls. Because it is likely that the pathogenicity and penetrance of CNVs at this locus vary, depending on the functional effect they have on the gene, we next restricted the association analysis to those CNVs which disrupt exons. We found seven in cases (0.17%) and five in controls (0.020%); these were more strongly associated with schizophrenia and had a higher odds ratio (P= 0.0027; OR 8.97, 95% CI 1.8-51.9).

Because almost half of the cases (but not controls) with CNVs were from Germany (6/14), we analysed haplotypes in the 5' region of the *NRXN1* gene region to see if we could detect evidence of a common ancestor for these six CNVs; however there was no evidence for haplotype sharing in the region, unlike the haplotype sharing observed for the common small deletion, supporting the notion that the majority of CNVs at this locus are not derived from a founder.

There was no sex difference among carrier subjects; CNVs were found in eight female and six male schizophrenia subjects, 31 female and 21 male controls. CNV type, geographic origin, SNP boundaries, position (Human genome build 36) size, phenotype and sex are shown in supplementary table 2. In addition to the deletions in patients with schizophrenia and control subjects with no known mental illness, there were five further deletions found in the Icelandic population, in a subject with autism, three subject with alcohol dependence and one first degree relative of a schizophrenia subject who did not participate in this study. These subjects are shown in figure 1 but were not included in the statistical analysis.

The clinical characteristics of the deletion subjects are shown in table 1; there seems to be nothing remarkable about these cases. We also tested allelic association for all SNP markers in the 2p16.3 region studied here, but no marker remained significantly associated at the 5% level after correction for number of performed tests (data not shown).

To address the issue of the role of splicing in *NRXN1* deletions, we reviewed all available genomic data across the *NRXN1* locus for evidence of alternative NRXN1 isoforms and identified EST and cDNA transcript evidence to support two additional isoforms, which share many exons with NRXN1a, which we term NRXN1a-2 and NRXN1 a-3 (figure 3). The NRXN1a-2 isoform shares exons 6-14 with NRXN1a, but employs alternative first and last exons. These alternative exons allow for distinct expression analysis of the NRXN1a-2 isoform, using probe 1558708_at on the Affymetrix HG-U133_Plus_2.0 GeneChip (supplementary figure 1c). This shows very low expression of NRXN1a-2, predominantly in the brain, compared to much higher brain specific expression of NRXN1a and NRXN1b (supplementary figures 1a & 1b). The NRXN1a-2 isoform has no signal peptide and no transmembrane domain, however it does contain Laminin G domains 2-4 and EGF domains 2 and 3 (figure 3). Evidence for the NRXN1a-3 isoform is much less conclusive as it shares all exons with exons 4-18 of NRXN1a.

DISCUSSION

In the present study, we used 2,977 schizophrenia patients and 33,746 controls, from seven European populations, to examine CNVs at the neurexin 1 locus on 2p16.3. Deletions at this locus have previously been associated with autism and mental retardation (6, 15-19), and deletions were also found in two families with schizophrenia. In addition we examined the other two known neurexin genes, *NRXN2* (11p13.1) and *NRXN3* (14q31.1) to see if they also harbour potentially pathogenic deletions. Kirov et al. (9) screened 93 cases, and found a deletion at 2p16.3 in one affected proband, as well as the unaffected mother and an affected sibling. Walsh et al. (11) found a different deletion disrupting *NRXN1* in identical twins concordant for childhood-onset schizophrenia. However because of the small sample size employed it is not possible to determine extent of association and the prevalence and odds ratio for disease risk from their data. In our study population, we found an excess of CNVs in the *NRXN1* gene in schizophrenia, 0.47% of cases compared to 0.15% of controls, which did not reach statistical significance. Conditioning on exon disrupting deletions or duplications (i.e. those more likely to be pathogenic) the association became significant ($P = 0.0027$; OR 8.97, 95% CI 1.8-51.9) indicating that CNVs in the region may have different pathogenicity or penetrance. There were no CNVs detected in the *NRXN2* or *NRXN3* loci, indicating that deletion or duplication of either of these genes is not a significant factor in disease.

De novo versus segregating deletions

One of the 2p16.3 deletions reported here (autistic subject, supplementary table 2), was also found *de novo* in the Icelandic population (12), and the remaining rare deletions probably occur *de novo* because they do not show any evidence of a founder effect and the boundaries differ between individuals. However we cannot be sure of this as in the present study, the majority of cases with deletions do not have parents and other relatives available for analysis. In contrast, the smallest common deletion appears to be a segregating polymorphism with a single founder. In the Kirov et

al. (9) study, the *NRXN1* deletion they found was scored as a segregating polymorphism as it was present in a mother and son; in the Walsh et al. (11) study it was found in a pair of MZ twins, with the parents not tested, so it is unclear if it arose by a *de novo* event. These deletions may still be predominantly *de novo* as in the cases found by Kirov et al. (9) the deletion in the mother could have occurred *de novo* in a transmission from a grandparent and subsequently been transmitted to her offspring; strong negative selection pressure (the reduced fecundity in male schizophrenics) would then mean that this deletion is quickly extinguished from the population. Reduced fecundity is well established in schizophrenia, and earlier onset (such as appears to occur in many cases with large CNVs) is also associated with lower fecundity. Although this is seen in both males and females, it is more pronounced in males (20-22). As Bassett et al. (20) suggest, these epidemiological observations are consistent with “a high mutation rate, consistent with unstable mutations, to explain the high prevalence of schizophrenia.”

Heterogeneity of deletions and potential impact on the complex splicing of neurexin

The heterogeneity of the deletions we have identified at the *NRXN1* locus also presents difficulties when it comes to assessing association with schizophrenia. Caution is needed in interpreting these results as the association we found was not significant when all CNVs were examined collectively, and required conditioning on exon-disrupting deletions. It is likely that pathogenicity of the deletions and duplications at this locus will vary, depending on the functional importance of the DNA deleted. Conditioning on exons is the simplest method of addressing this, as protein structure and splicing will be affected, and under this test the association between exonic deletions and schizophrenia in our sample is highly significant.

The complexity of neurexin regulation at the level of alternative splicing has already been suggested, with six splice sites with perhaps 1000 isoforms (27) and the potential to theoretically generate over 2000 isoforms encoding variant extracellular domains (23); . Although the exact impact of deletions of the first exons of *NRXN1a* on alternative splicing is unclear, it is possible

that the expression of alternative NRXN1 isoforms may compensate for the loss of the NRXN1a transcript. It is interesting that the first exon of NRXN1a-2 seems to be deleted in the small founder deletion observed in 24 subjects most of whom were controls. Considering the very low levels of expression of NRXN1a-2, and its frequent deletion in controls, we therefore suggest that NRXN1a-2 may have a very limited biological role in Neurexin signalling. However, activity of the NRXN1a-2 isoform may become significant against a background of deletion of the other Neurexin isoforms. In such circumstances, even low-level dominant negative activity may become biologically relevant. As most known Neurexin interactions are extracellular it is unclear if NRXN1a-2 has any function, although if it is secreted it may exert a dominant negative effect by competitive binding with NRXN1a interactors; alternatively improper cellular localization of NRXN1a-2 may also be deleterious.

Evidence for the NRXN1a-3 isoform is much less conclusive than other isoforms as it shares all exons with exons 4-18 of NRXN1a. This excludes the possibility of differentiation of NRXN1a and NRXN1a-3 by expression analysis. There is a high level of mammalian conservation before the first exon of NRXN1a-3 and there is also EST evidence to support the existence of this isoform although, these ESTs cannot be clearly differentiated from NRXN1a transcripts. The putative initiation methionine of the NRXN1a-3 isoform correlates with NRXN1a Met321 roughly a third of the way into Laminin G domain 2 of the protein (see figure 3).

The truncated domain of NRXN1a-3 may not be functional as it has lost a critical Ca^{2+} binding residue Asp306, which also mediates interactions with Neurexophilins (24, 25). Loss of Neurexophilin interactions may be particularly relevant to schizophrenia pathology, as Neurexophilin-3 knockout mice show no anatomical defects, but remarkable functional abnormalities in sensory information processing and motor coordination, evident by increased startle response, reduced prepulse inhibition, and poor Rotarod performance (26). It is tempting to speculate that the disruption of NRXN1/Neurexophilin interactions that are likely to be seen in all our observed exonic deletions may explain some of the pathology of schizophrenia. More detailed

bioinformatic and functional biology may be able to shed further light on the genotype-phenotype and functional consequences of various deletions at this locus.

Haploinsufficiency as a mechanism of pathogenicity

In addition to the disruption of splicing and the effects of this on the interaction of neurexin proteins with protein partners, the association between *NRXN1* deletions and schizophrenia could also result from haploinsufficiency of the gene. Given that we see a strong association when conditioning on disruption of an exon in the two known RefSeq sequences for *NRXN1*, and the most common and smallest deletion of intron only sequence (although the first exon of the novel splice variant *NRXN1a-2* (figure 3) appears to be deleted) shows a founder effect and is thus not likely to be under strong negative selection pressure, it is not likely that all the deletions we found at the *NRXN1* locus have the same biological effect on disrupting the function of *NRXN1*, and some will probably be neutral variants. Intron-only deletions could of course disrupt splicing or remove essential regulatory elements such as splicing enhancers and/or silencers, and the neurexin 1 gene is extensively spliced, with more than 2000 potential isoforms (27). In order to provide further information on the effects of these deletions it will be necessary to analyse mRNA and protein expression, for example using relative allelic expression of non-deleted exons in *NRXN1* mRNA to demonstrate which deletions result in haploinsufficiency.

Founder effects and the distribution of neurexin 1 deletions in Europe

There is also a question of how to treat the sample with regard to geographical origin, as most of the deletions in cases were from the German population (6/14; 43%) although of course as 25% of the examined patients were German in origin, this over-representation could have occurred by chance. If these deletions are truly recurrent, and depend on a random mechanism involving DNA sequences or recombination rates in the region, one would not expect significant geographic clustering. Clustering may indicate deletions which are segregating in the population, perhaps from

an ancestral founder, or a pre-mutation or polymorphic DNA sequence which increases the probability of a *de novo* event and is more prevalent in one population than others. A haplotype analysis that we performed on the German samples with deletions showed no excess sharing of alleles around the *NRXN1* 5' region, which, when combined with the observed heterogeneity of the deletions, strongly opposes a founder effect. The most common deletion, which occurred in 29 subjects, was also the smallest and showed evidence of a single founder. This may be a neutral segregating polymorphism which does not significantly affect *NRXN1* function and is not relevant to risk of schizophrenia.

Penetrance and frequency of neurexin deletions

Rare *NRXN1* deletions have also been reported in the normal population. Pinto et al., (28) found an intronic 71kb deletion in a sample of 506 unrelated healthy individuals from Northern Germany and 270 HapMap individuals (chr2:50,826,443-50,897,556) which overlaps with some of the intronic deletions seen in the present study, and both Redon et al., (29) (chr2:50,643,118-50,686,339 Build 36) and Wong et al. (30) found similar sized deletions (chr2:50,651,965-50,714,553) in single control subjects, one of which removes four exons of the gene. These subjects have no known neuropsychiatric phenotype, consistent with the notion that although deletions of *NRXN1* are strongly associated with mental illness they are not necessarily fully penetrant, as seen in the data from the present study, and some individuals may be able to compensate for dosage loss, for example through a high expressing residual copy of the gene, another modifying locus, or an environmental event which results in a gene-by-environment interaction pathway to disease.

The 2p16.3 deletions we found are moderately rare, occurring at a rate of about 4.7 per 1000 cases, and, from the control population in the present study, at about 0.15% of live births. This is substantially higher than the VCFS deletion, found in about 0.01% of live births (31), making it one of the most common gene deletions known. However if we exclude the deletions in *NRXN1* which do not disrupt exons and are less likely to be pathogenic, the deletion rate is 1.7 per 1000 (0.17%) in

the cases compared to 0.020% in the controls, a rate very comparable to VCFS deletions.

The structural mutations showing association to schizophrenia disrupt the *NRXN1* gene, rather than deleting or duplicating the entire gene. A generalised study of structural mutations by Lupski and Stankiewicz (32) showed that such mutations are much more likely to have biological consequences, in our case deletion of exons from the most prevalent NRXN1a isoform, may favour the expression of the rarer NRXN1a-2 and NRXN1a-3 isoforms, which may possibly exert a dominant negative effect on Neurexin signalling, particularly through Neurexophilins.

The odds ratio (~10 fold increase in risk) is less than that associated with the VCFS deletion, where about 25% of subjects develop psychosis and the odds ratio for the VCFS deletion is about 30, assuming a population rate of non-VCFS deletion schizophrenia of about 1% (33). It is possible that there are many more CNVs in the human genome which can cause schizophrenia, and these could make up a significant proportion of genetic risk for the disease which could prove clinically useful, for example through genetic testing.

Potential deletion mechanisms at the NRXN1 locus

The nature of the loci studied here raises questions about the mechanisms underlying CNV events. Many chromosomal rearrangements are not random events, but the result of genome architecture that may create inherent instability (34), such as the long low copy repeats (LCRs) typical of many other regions where CNVs have been reported. These result in non-allelic homologous recombination (NAHR) between the LCRs on opposite flanks of the deleted/duplicated segment. However several genomic disorders are associated with rearrangements whose breakpoints do not cluster within LCRs, as in the case of *NRXN1* deletions. There we see no evidence of LCRs flanking the breakpoints, nor do we find common breakpoints or evidence of sequence duplications in the region. The LD pattern is particularly unstructured in the 5' end of the gene and the upstream region where we observe the deletions. This may suggest increased recombination rate in the region that may be related to the aggregation of deletions at the site, through an unknown mechanism. There are suggestions that CNV events are also associated with the presence of processed

pseudogenes, the presence of microsatellites, and non-homologous end joining mediated by micro-homologies (<25bp homology), especially in subtelomeric regions, or DNA replication-based mechanisms, rather than a NAHR mechanism mediated by larger matching repeats (32, 35, 36). Similarly to *NRXN1*, copy number variants in the *PLP1* gene do not show common breakpoints. Lee et al. (36) examined this locus and uncovered evidence for sequence complexity at some junctions, consistent with a DNA replication-based mechanism of CNV formation (replication Fork Stalling and Template Switching, FoSTeS). This mechanism might explain the non-recurrent apparently *de novo* deletions with multiple breakpoints seen in *NRXN1*. To examine this further, we performed long range phasing analysis (37) of the single case of autism known to have a *de novo NRXN1* deletion in the Icelandic population. This showed that no recombinations had occurred in the region around the *de novo* deletion (data not shown), providing further evidence against a role for an NAHR mechanism in this case.

Schizophrenia and autism: aetiological overlap?

The data from our sample of cases and controls provides support for a role for deletions within the *NRXN1* gene in schizophrenia, and raises the concept that this disorder has shared genetic risk factors with autism. If this is indeed the case, it would change the way the distinctions between the diseases is formulated, and therefore has profound clinical implications. In support of this are previous observations such as the VCFS deletions at 22q11.2, which appear to increase the risk of both autism (38) and schizophrenia (33, 39), and data from the present and other studies indicating that neurexin deletions are seen in both disorders as well as other behavioural syndromes (9, 11, 15-19; 40). On the other hand, the concept of diagnostic overlap between the two disorders is controversial: while Rzhetsky et al. (42) provides some theoretical support for aetiological overlap, evidence from clinical and epidemiological study is less supportive (42,43). In general there have been few genetic and epidemiological studies of potential overlap between the disorders, and the present and other studies point the way towards more comprehensive evaluation of potential diagnostic overlap between these and indeed other psychiatric disorders in the future.

MATERIALS AND METHODS

Samples

A total of 2977 affected and 33746 controls from six European populations were successfully examined for CNVs at the *NRXN1*, *NRXN2* and *NRXN3* loci studied here (supplementary table 1); 1435 schizophrenia patients and 28554 control individuals from the Iceland, Scotland, Germany, England, Italy and Finland (The SGENE sample; <http://www.SGENE.eu>), 491 affected and 881 controls from Bonn, Germany, 806 cases and 4039 controls from the Netherlands and 245 cases and 272 controls from Norway. For a full description of samples see Supplementary Information. Ethical approval was obtained from the local Ethics Committees. All participants gave written informed consent.

The SGENE samples were typed on the HumanHap300 BeadArrayTM (Illumina, San Diego, USA) at deCODE genetics. The samples from Bonn were typed at Bonn University on HumanHap550v3 BeadArrayTM (Illumina, San Diego, USA). The Dutch samples from Utrecht University were genotyped at the University of California, Los Angeles, on HumanHap550v3 BeadArrayTM (Illumina, San Diego, USA). The remaining Dutch samples were genotyped at deCODE genetics on HumanHap300 BeadArrayTM (Illumina, San Diego, USA). The Norwegian samples were genotyped on Affymetrix GeneChip GenomeWide SNP 6.0 array and analyzed using the Affymetrix Power Tools 1.8.0. Samples with Contrast QC below 0.4 were excluded as recommended by the manufacturer.

CNV detection

DosageMiner software, developed at deCODE genetics, and QuantiSNP (44; see below for description) was used to identify deletions and duplications around the regions reported by Kirov et al. (9) and Walsh et al. (11) in all samples except the Norwegian sample, where only QuantiSNP was used. DosageMiner, described in detail elsewhere (12), uses the intensities from SNP probes on the Illumina microarrays to estimate copy number of genomic regions, and models factors such as SNP effect, sample effect and GC-content in the neighbouring region to normalise the intensities.

The software then automatically registers SNP loci where intensities fall above or below an empirical threshold. We conditioned our search to include only CNVs with 10 or more contiguous markers registered as duplicated or deleted by DosageMiner, in order to reduce the level of false positive CNVs identified by the software. We then inspected the intensities by eye for all samples with a CNV identified by the software. We used SNP genotypes for further confirmation in case of deletions by conditioning on homozygosity for all markers within the boundaries of the suspected CNV. In some instances, the actual CNV may be shorter than 10 contiguous markers, because the boundaries called by DosageMiner may not always be entirely accurate.

The QuantiSNP software, developed at the Wellcome Trust Centre for Human Genetics and at the University of Oxford (<http://www.well.ox.ac.uk/QuantiSNP/>), relies on an Objective Bayes Hidden-Markov Model to estimate copy number variations (44). In this model, the hidden states denote the unknown copy number at the inspected SNPs. Genotype data was used to compute different states. Based on the ratio of fluorescent dye ratios (logR) and stretches of homozygosity, the algorithm computes a Bayes factor that is used to calibrate the model to a fixed type I (false-positive) error rate. A Bayes factor threshold of 10 is considered as a promising value for the possible presence of a CNV. Usually, such values occur when 5-10 consecutive SNPs are deleted/duplicated. Differences in GC base pairs may result in biased hybridization behaviour of SNP probes bearing the risk of miscalling genotypes. To normalize for this, QuantiSNP assigns a locus-specific GC value to each probe. Again, all potential CNVs detected were subsequently visually inspected and confirmed using the BeadStudio software (Illumina, San Diego, USA). We used the Cochran-Mantel-Haenszel exact test for the association analysis, which includes a correction for population of origin in the analysis. In order to condition on exons, we scored a CNV as exon-disruption if it overlapped any known exonic sequence in NRXN1, based on reviewed sequence from Human mRNA from RefSeq (NM_004801 for NRXN1-alpha). Long-range phasing analysis to detect recombination events in the Icelandic case of autism with a *de novo* NRXN1 deletion was performed using the method of Kong et al. (37).

ACKNOWLEDGEMENTS

We want to thank the subjects and their relatives and staff at the recruitment centres. This work was sponsored by EU grant LSHM-CT-2006-037761 (Project SGENE).

CONFLICTS OF INTEREST

No reported conflict of interests.

REFERENCES

1. Sebat, J., B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, B. Yamrom, S. Yoon, A. Krasnitz and J. Kendall et al., (2007) Strong association of de novo copy number mutations with autism. *Science*, **316**, 445–449.

2. Zhao, X., A. Leotta, V. Kustanovich, C., Lajonchere, D.H., Geschwind, K., Law, P. , Law, S., Qiu, C., Lord, J. Sebat, J., et al., (2007) A unified genetic theory for sporadic and inherited autism, *Proc. Natl. Acad. Sci. USA*, **104**, 12831–12836.

3. Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson H., Ferreira M.A., Green T., et al., (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667-675.

4. Reichenberg, A., Gross, R., Weiser, M., Bresnahan, M., Silverman, J., Harlap, S., Rabinowitz, J., Shulman, C., Malaspina, D., Lubin, G., Knobler, H.Y., Davidson, M., Susser E. (2006) Advancing paternal age and autism. *Arch. Gen. Psychiatry.* **63**, 1026-1032.

5. Malaspina, D., Harlap, S., Fennig, S., Heiman, D., Nahon, D., Feldman, D., Susser, E.S. (2001) Advancing paternal age and the risk of schizophrenia. *Arch. Gen. Psychiatry*, **58**, 361-367.

6. Bish J.P., Bearden C.E., Ding L., Ferrante S., Nguyen V., Gee J.C., McDonald-McGinn D.M., Zackai E.H., Emanuel B.S. (2005) A multilevel analysis of cognitive dysfunction and psychopathology associated with chromosome 22q11.2 deletion syndrome in children. *Dev. Psychopathol.* **17**, 753-784.

7. Millar, J.K., Wilson-Annan, J.C., Anderson, S., Christie, S., Taylor, M.S., Semple, C.A., Devon, R.S., St Clair, D.M., Muir, W.J., Blackwood, D.H., Porteous, D.J. (2000) Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum. Mol. Genet.* **9**, 1415-1423.

8. Murphy, K.C. (2002) Schizophrenia and velo-cardio-facial syndrome. *Lancet* **359**, 426-430.

9. Kirov, G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., O'Donovan, M.C., Erdogan, F., Owen, M.J., Ropers, H.H., Ullmann, R. (2008) Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum. Mol. Genet.*, **17**, 458-465.

10. Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., Karayiorgou, M. (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet.* **40**, 880-885.

11. Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., et al., (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539-543.

12. Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O.P.H., Ingason, A., Fossdal, R., Steinberg, S., Sigurdsson E., Sigmundsson T., Hansen T., et al. (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232-236.

13. The International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237-241.

14. Mizuguchi T, Hashimoto R, Itokawa M, Sano A, Shimokawa O, Yoshimura Y, Harada N,

- Miyake N, Nishimura A, Saitsu H, Sosonkina N, Niikawa N, Kunugi H, Matsumoto N. (2008) Microarray comparative genomic hybridization analysis of 59 patients with schizophrenia. *J. Hum. Genet.* **53**, 914-919
15. Feng, J., Schroer, R., Yan, J., Song, W., Yang, C., Bockholt, A., Cook, E.H. Jr., Skinner, C., Schwartz, C.E., and Sommer, S.S. (2006). High frequency of neurexin 1beta signal peptide structural variants in patients with autism. *Neurosci. Lett.*, **409**, 10–13.
16. Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.Q., Vincent, J.B., Skaug, J.L., Thompson, A.P., Senman, L., et al. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.* **39**, 319–328.
17. Marshall C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008) Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477-488.
18. The Autism Genome Project Consortium. (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.*, **39**, 319–328.
19. Kim, H.G., Kishikawa, S., Higgins, A.W., Seong, I.S., Donovan, D.J., Shen, Y., Lally, E., Weiss, L.A., Najm, J., Kutsche, K., et al. (2008) Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.*, **82**, 199-207.
20. Bassett, A.S., Chow, E.W., Weksberg, R., Brzustowicz, L. (2002) Schizophrenia and genetics: new insights. *Curr. Psychiatry Rep.*, **4**, 307-314

21. Haukka, J., Suvisaari, J., Lönqvist, J. (2003) Fertility of patients with schizophrenia, their siblings, and the general population: a cohort study from 1950 to 1959 in Finland. *Am. J. Psychiatry*, **160**, 460-463.

22. Terzian, A.C., Andreoli, S.B., Razzouk, D., Chaves, A.C., Mari Jde, J. (2006) Fertility and fecundity of an outpatient sample with schizophrenia. *Rev. Bras. Psiquiatr.*, **28**, 305-307.

23. Rowen, L., Young, J., Birditt, B., Kaur, A., Madan, A., Philipps, D.L., Qin, S., Minx, P., Wilson, R.K., Hood, L., Graveley, B.R. (2002) Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. *Genomics*, **79**, 587-597.

24. Sheckler, L.R., Henry, L., Sugita, S., Südhof, T.C., Rudenko, G. (2006) Crystal structure of the second LNS/LG domain from neurexin 1alpha: Ca²⁺ binding and the effects of alternative splicing. *J Biol Chem.* **281**, 22896-22905.

25. Missler, M., Hammer, R.E., Südhof, T.C. (1998) Neurexophilin binding to alpha-neurexins. A single LNS domain functions as an independently folding ligand-binding unit. *J Biol Chem.*, **273**, 34716-34723.

26. Beglopoulos, V., Montag-Sallaz, M., Rohlmann, A., Piechotta, K., Ahmad, M., Montag, D., Missler, M. (2005) Neurexophilin 3 is highly localized in cortical and cerebellar regions and is functionally important for sensorimotor gating and motor coordination. *Mol. Cell. Biol.*, **25**, 7278-7288.

27. Ullrich, B., Ushkaryov, Y.A., Südhof, T.C. (1995) Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron*, **14**,

497-507.

28. Pinto, D., Marshall, C., Feuk, L., Scherer, S.W. (2007) Copy-number variation in control population cohorts. *Hum. Mol. Genet.*, **16** Spec No. 2:R168-73.
29. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006) Global variation in copy number in the human genome. *Nature*, **444**, 444-454.
30. Wong, K.K., deLeeuw, R.J., Dosanjh, N.S., Kimm, L.R., Cheng, Z., Horsman, D.E., MacAulay, C., Ng, R.T., Brown, C.J., Eichler, E.E., Lam, W.L. (2007) A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.* **80**,
31. Tezenas Du Montcel, S., Mendizabai, H., Ayme, S., Levy, A., Philip, N. (1996) Prevalence of 22q11 microdeletion. *J. Med. Genet.*, **33**: 719.
32. Lupski, J.R., Stankiewicz, P. (2005) Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet.* **1**, e49.
33. Murphy, K.C., Jones, L.A., Owen, M.J. (1999) High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch. Gen. Psychiatry* **56**, 940-945.
34. Shaw, C.J., Lupski, J.R. (2004) Implications of human genome architecture for rearrangement-based disorders: The genomic basis of disease. *Hum. Mol. Genet.*, **13**, Spec No 1:R57–64.
35. Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ. (2005) Human

subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* 437, 94-100.

36. Lee, J.A., Carvalho, C.M., Lupski, J.R. (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, **131**, 1235-1247

37. Kong A., Masson G., Frigge M.L., Gylfason A., Zusmanovich P., Thorleifsson G., Olason P.I., Ingason A., Steinberg S., Rafnar T., Sulem P., Mouy M., Jonsson F., Thorsteinsdottir U., Gudbjartsson D.F., Stefansson H., Stefansson K. (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068-1075.

38. Antshel, K.M., Aneja, A., Strunge, L., Peebles, J., Fremont, W.P., Stallone, K., Abdulsabur, N., Higgins, A.M., Shprintzen, R.J., Kates, W.R. (2007) Autistic spectrum disorders in velo-cardio facial syndrome (22q11.2 deletion). *J. Autism Dev. Disord.* **37**, 1776-1786.

39. Chow, E.W., Bassett, A.S., Weksberg, R. (1994) Velo-cardio-facial syndrome and psychotic disorders: implications for psychiatric genetics. *Am. J. Med. Genet.* **54**, 107-112.

40. Zahir, F.R., Baross, A., Delaney, A.D., Eydoux, P., Fernandes, N.D., Pugh, T., Marra, M.M., Friedman, J.M. (2007) A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1{alpha} *J. Med. Genet.* **45**, 239-243.

41. Rzhetsky, A., Wajngurt, D., Park, N., Zheng, T. (2007) Probing genetic overlap among complex human phenotypes. *Proc. Natl. Acad. Sci. U S A.*, **104**, 11694-11699.

42. Volkmar, F.R., Cohen, D.J. (1991) Comorbid association of autism and schizophrenia. *Am. J. Psychiatry*, **148**, 1705-1707.
43. Esterberg, M.L., Trotman, H.D., Brasfield, J.L., Compton, M.T., Walker, E.F. (2008) Childhood and current autistic features in adolescents with schizotypal personality disorder. *Schizophr. Res.* **104**, 265-273.
44. Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., Ragoussis J. (2007) QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.*, **35**, 2013-2025.

LEGENDS TO FIGURE

Figure 1:

UCSC browser output showing the positions of the 2p16.3 CNVs from the Kirov et al., (2008), Walsh et al., (2008), Friedman et al., 2008 and Szatzmar et al., 2008 studies (blue lines) relative to the NRXN1 gene, and the CNVs discovered in the present study (schizophrenia, red lines; other psychiatric diagnoses, brown lines; controls, green lines). The majority of the discovered CNVs are deletions, asterisks indicate duplications. Markers from the Illumina 300K and 550K arrays, segmental duplications of >1000 bp as well as LD structure of the Hapmap CEU sample (r^2) is also shown.

Figure 2:

UCSC browser output showing the positions of exon-disrupting CNVs discovered in the study relative to the 2p16.3 CNVs from the Kirov et al., (2008), Walsh et al., (2008), Friedman et al., 2008 and Szatzmar et al., 2008 studies and known (schizophrenia, red lines; other psychiatric diagnoses, brown lines; controls, green lines; previously described CNVs, blue lines). The four putative Neurexin isoforms are shown below the deletions, along with protein domains aligned to genomic sequence.

Figure 3:

Protein domain organisation of Neurexin 1 isoforms. NRXN1a contains an N-terminal signal peptide, six laminin G (LamG) domains, three epidermal growth factor-like (EGF) domains, a transmembrane domain and a short cytoplasmic domain. NRXN1b contains a different signal peptide, but shares the last LamG domain, transmembrane domain and cytoplasmic domain with NRXN1a. NRXN1a-2 shares the 2nd, 3rd, 4th and 5th LamG domains and the 2nd and 3rd EGF domains with NRXN1a, but does not contain a signal peptide or transmembrane domain. NRXN1a-3 has no signal peptide and has a truncated version of the 2nd LamG domain but shares all remaining domains with NRXN1a. The five regions in the NRXN1 gene where alternative splicing occurs, leading to insertion or deletion of amino acids is indicated by arrows (SS#1-5). Protein domain annotation was generated using SMART (<http://smart.embl-heidelberg.de/>), using swissprot accessions Q9ULB1 (NRXN1a) and P58400 (NRXN1b) and translations from the genbank transcript AK093260 (NRXN1a-2) and Ensembl transcript ENST00000331040 (NRXN1a-3).

Table 1 Phenotypic features of schizophrenia cases with 2p deletions.

Centre	Diagnosis	Family history of schizophrenia	CNV boundaries	Age at onset	Sex	Other
Munich 1	DSMIV:295.3	no	chr2:50,856,110-50,900,862	27	F	.
Munich 2 ¹	DSMIV:295.3	no	chr2:50,890,216-51,116,653	25	M	Social contact problems in childhood
Munich 3	DSMIV:295.3	no	chr2:51,147,600-51,225,851	31	F	.
Verona 1 ^{1, 2}	ICD10: F20.3	Brother with mental illness (not specified)	chr2:50,071,499-50,208,992	36	F	Chronic, positive symptoms, low IQ (82), low educational level
Verona 2	ICD10: F20.3	Yes other psychiatric disorder	chr2:50,822,312-50,948,557	14	F	Chronic, negative symptoms, episode of aggression
Helsinki 1 ¹	DSMIV:295.3	Yes, other psychiatric disorder	chr2:51,101,161-51,344,213	37	M	Alcoholism
Utrecht 1	DSMIV:295.3	Probably psychosis in first or second degree relative	chr2:50,735,657-50,800,548	29	F	Not mentioned
Utrecht 2 ²	DSMIV:295.9	Not examined	chr2:50,786,446-50,900,862	19	M	Not mentioned
Utrecht 3 ¹	DSMIV:295.3		chr2:51,002,576-51,250,922	.	M	.
Utrecht 4 ¹	DSMIV:295.9	None	chr2:51,0249,62-51,251,873	Unknown	F	Episodic with partial remission and negative symptoms
Reykjavik 1	RDC:126.4	None	chr2:51,211,406-51,299,436	26	M	Not mentioned
Bonn 1 ¹	DSMIV:295.3	None	chr2:50,711,199-50,756,435	18	M	Myoclonic seizures in the right shoulder
Bonn 2	DSMIV:295.3	None	chr2:50,836,690-50,936,258	21	F	Resistncy in left lower abdomen with otherwise unremarkable finding
Bonn 3 ¹	DSMIV:295.3	Yes	chr2:50,850,456-51,225,851	27	F	Mental retardation (mild), Tardive dyskinesia.

¹exon disrupting²duplication

FIGURES

Figure 1

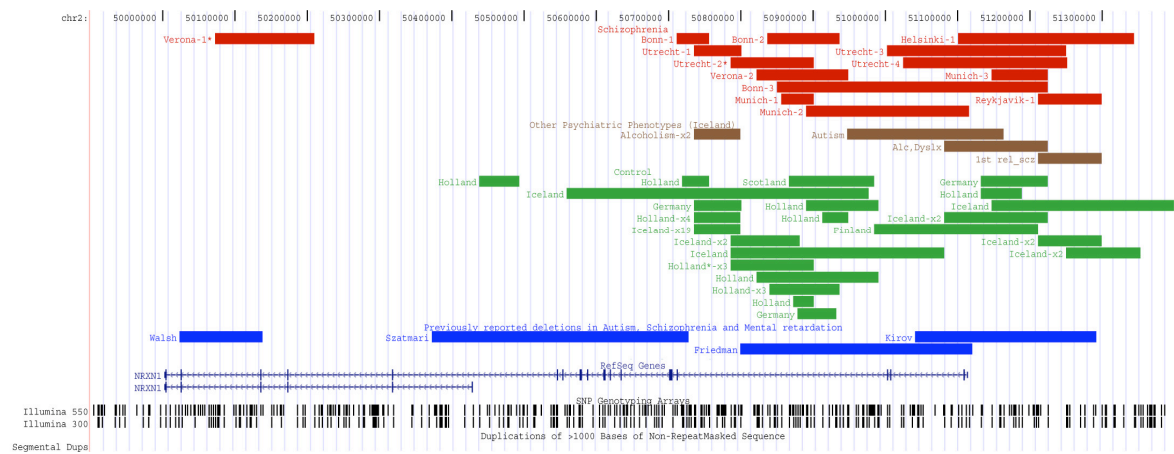


Figure 2

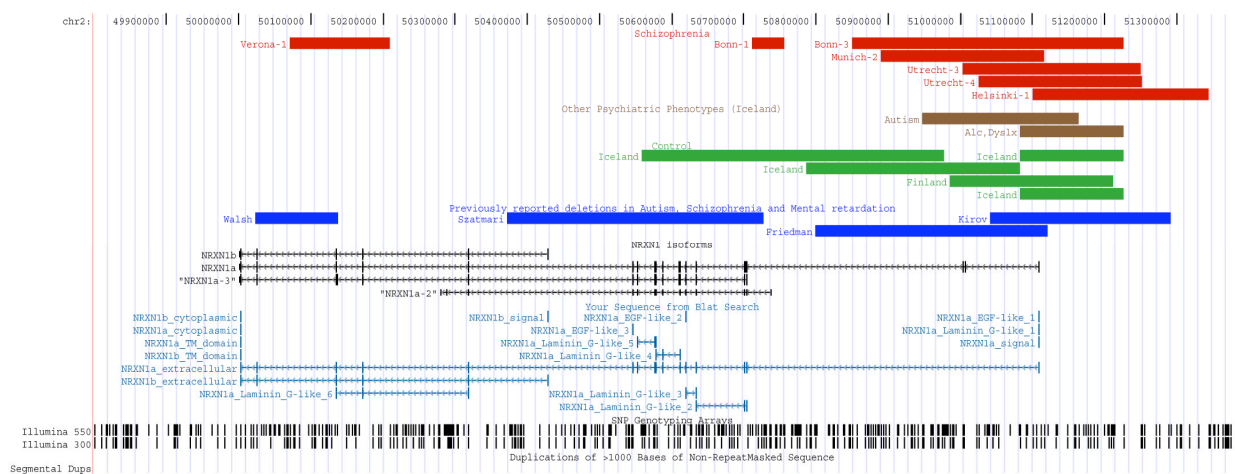
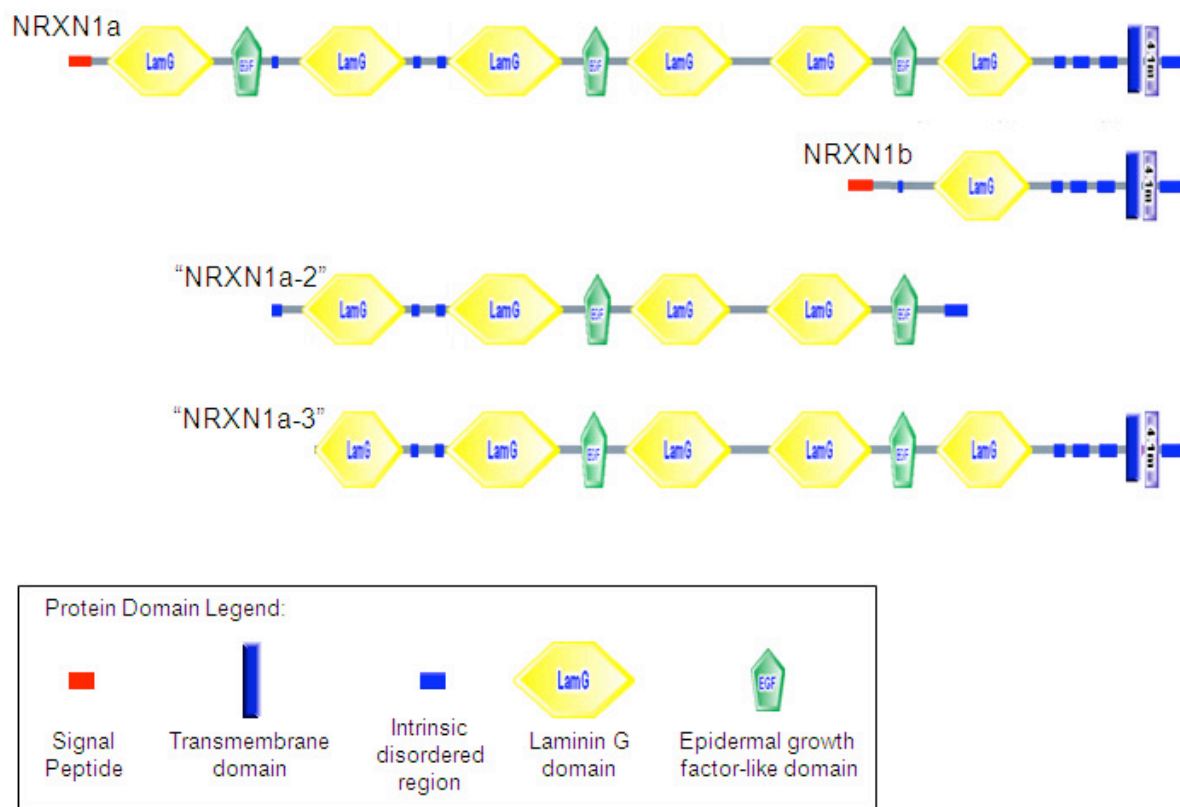


Figure 3

ABBREVIATIONS

CI	Confidence interval
CNV	Copy number variant
DISC1	Disrupted in schizophrenia 1
LCRs	low copy repeats
LD	linkage disequilibrium
NAHR	non-allelic homologous recombination
NRXN1	Neurexin 1
NRXN2	Neurexin 2
NRXN3	Neurexin 3
OR	Odds ratio
VCFS	velo-cardio-facial syndrome